EE698R: Advanced Topics in Machine Learning

Course Project

Project Title: Exploring Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN) for Image Restoration: A Comparative Study

Instructor: Prof. Vipul Arora

Student: Arindom Bora (Roll No : 210183)
Neerav Sethi (Roll No : 210661)

Date: 16/04/2024

Objective

1. Train a Variational Autoencoder (VAE) model and a Generative Adversarial Network (GAN) on a Image Restoration dataset that can generate clear images from noisy/damaged images.
2. Compare the outputs of the model with the ground truth using image similarity metrics.

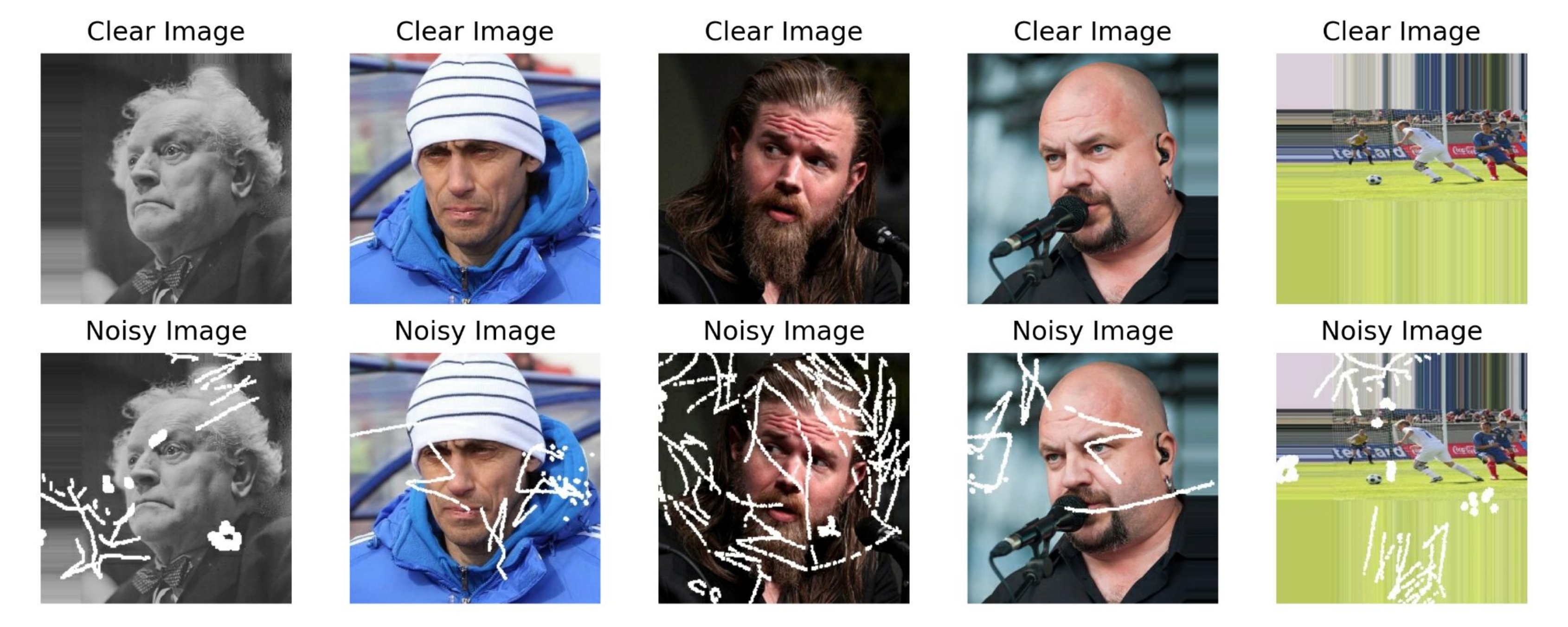
Dataset

Source: The dataset is sourced from Kaggle and is called “humanface8000”.

Size: It contains a total of 16000 images of human faces, 8000 clear images and 8000 noisy images.

Quality: The images are of high resolution and includes a diverse range of human faces with different backgrounds.

Preview of Dataset



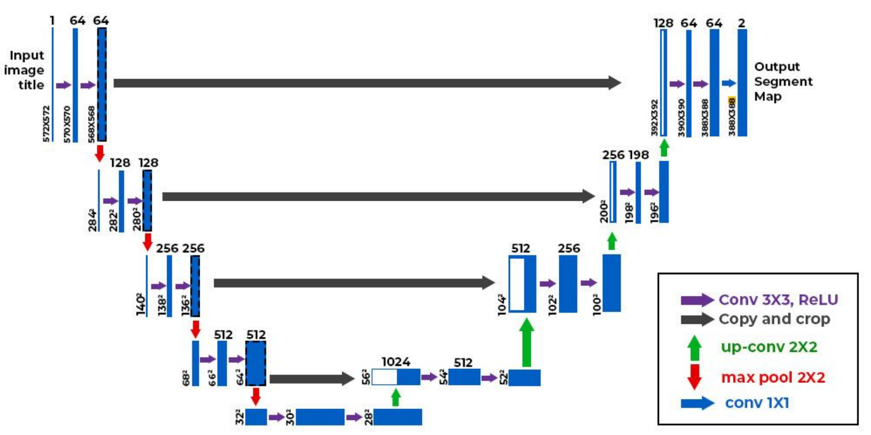
Data Preprocessing

1. **Reshaping :** The images in the dataset had various sizes. So they were initially reshaped to a uniform size of (256,256,3).
2. **Rescaling :** The pixel values initially ranged from 0 to 255. For efficient computation purposes, the pixel values are scaled down to the range (0,1).

U-Net Architecture

It is a widely used CNN based DL architecture that consists of a encoder layer and a decoder layer.

An important aspect of the architecture are the skip connections from encoder to decoder which helps to preserve spatial information and locate the features more accurately..

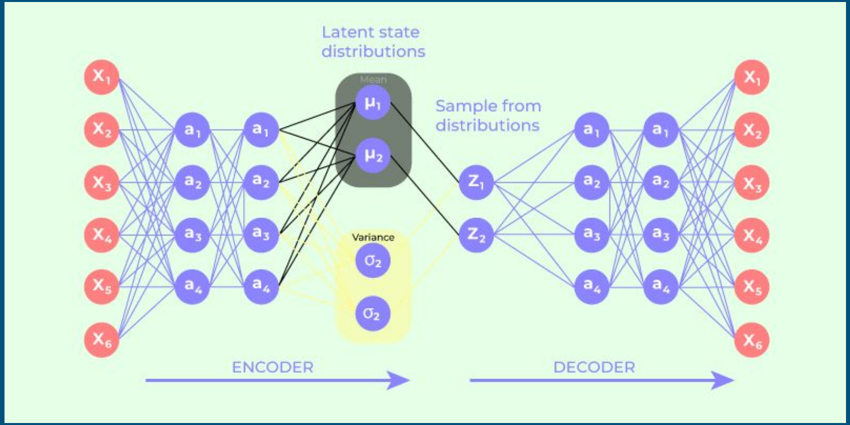


Variational Autoencoders (VAE)

It is a neural network architecture proposed in 2013 by Diederik P. Kingma and Max Welling that provides a probabilistic manner for describing an observation in latent space. It includes a encoder layer that outputs a probability distribution of the given data and a decoder layer that takes a sampled point from the latent distribution and reconstructs it back into data space.

Variational autoencoder utilises 2 loss functions:

1. Reconstruction loss: MSE between target image and generated image.
2. KL-divergence loss: Difference between the true probability distribution of target and learned probability distribution.

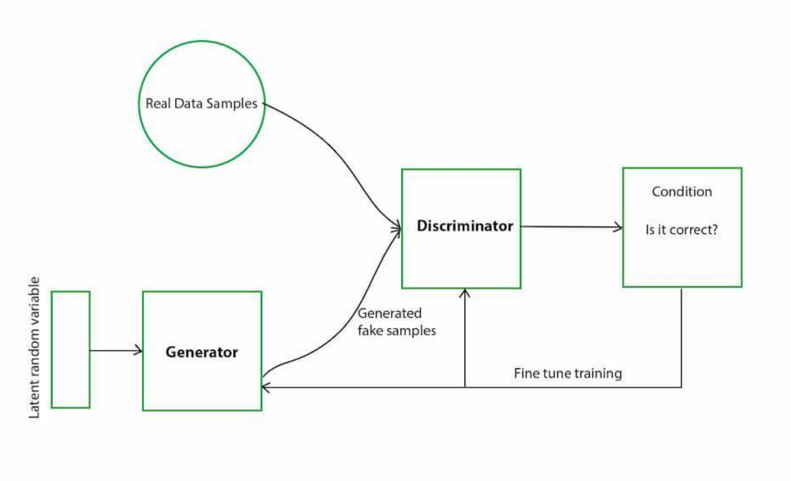


Generative Adversarial Networks (GAN)

It is a neural network architecture comprising of two models, a Generator and a Discriminator. The Discriminator attempts to accurately distinguish between generated and real data, while the Generator attempts to fool the Discriminator, by generating artificial data similar to real data. Realistic, high-quality samples are produced as a result of this adversarial training, which drives both networks toward advancement.

Deep Convolutional GAN (DCGAN) is one of the most popular implementations of GAN which comprises of multiple ConvNets followed by dense layers.

Max pooling layers in the ConvNets are replaced by convolutional stride.



Model Architecture and Training

1. **VAE :**

   a. **Encoder:** The encoder model of our VAE consists of 8 convolution layers. It takes in images of sizes (256,256,3) and encodes it into a 2 vectors, each of length 64, denoting the latent dimension of mean and variance. Each convolution layer is followed by batch normalization layer and a LeakyReLU activation layer. The outputs of each set of convolution layers are stored and later used in decoder.

   b. **Decoder:** The decoder model consists of 8 upsampling layers. It takes in encoded inputs of length 64, and returns images of shape (256,256,3). Each upsampling layer (except the last) is followed by batch normalization, ReLU activation and concatenation layers. The concatenation layers adds the encoder output of same size to the output of the upsampling layer. This helps in putting a constraint between model input and output.

The VAE model is trained for 100 epochs with adam optimizer of constant learning rate 0.0001.

2. GAN:

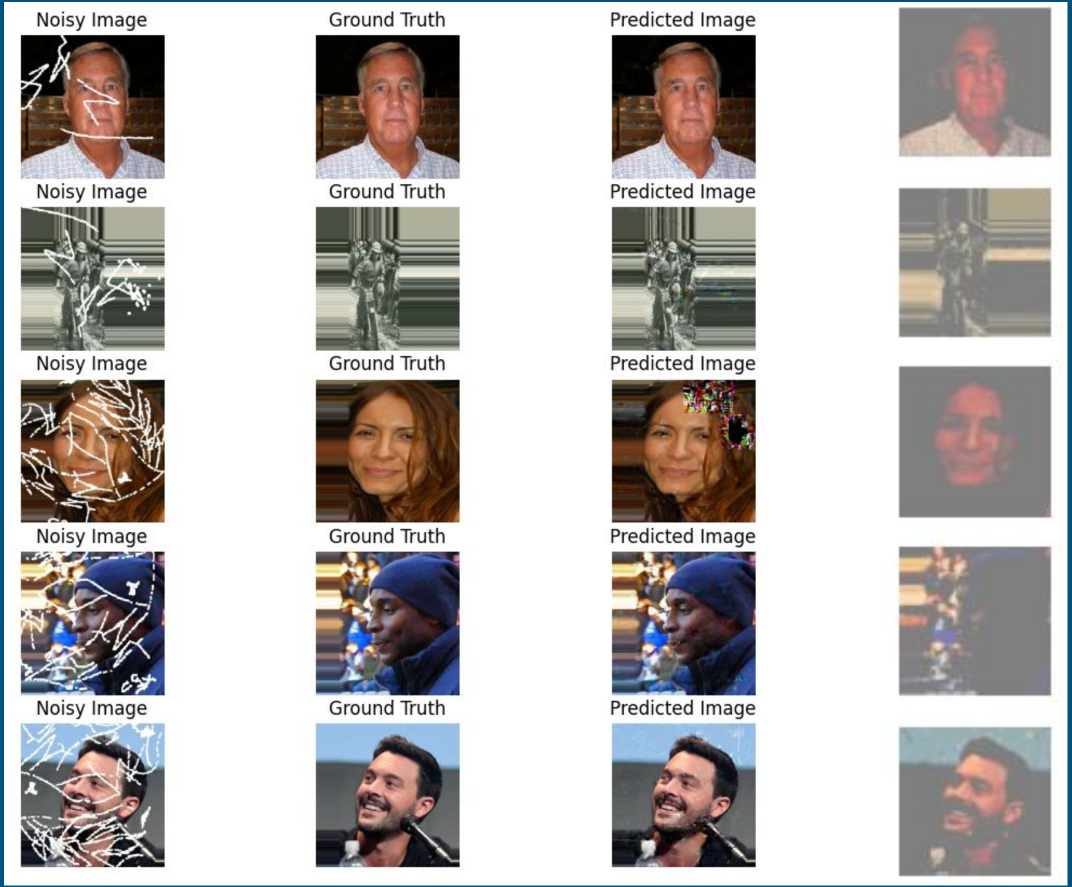
a. **Generator :** The Generator model consists of 8 downsampling layers that encodes input images to 512 length vector and 8 upsampling layers that generates output images from the encoded vector. As similar to the VAE decoder, the upsampling layers have concatenation layers that concatenates output of the downsampling layers at various stages.

b. **Discriminator :** The Discriminator model takes in 2 input images and contains 4 downsampling layers and final dense layers that outputs a single vector.

Both the generator and discriminator models are trained for 100 epochs with Adam optimizer and Binary Cross Entropy Loss function.

Results

Input Image Ground Truth GAN output VAE output



Discussion

We used 3 different metrics, namely, **Mean Squared Error, Normalized Root Mean Squared Error** and **Structural Similarity** to evaluate the images generated by the 2 models with the real images. The result are as follows:

Metric	**Noise Image**	**VAE generated image**	**GAN generated image**
Mean Squared Error	2522.92	4241.17	493.36
Normalized Root Mean Squared Error	0.405	0.536	0.161
Structural Similarity	0.754	0.399	0.844

Conclusions

1. GANs outperform VAE in Image Restoration task.
2. The outputs of VAE are blurred even with the same model architecture and training parameters. However, the white marks of the images are not seen on VAE generated images
3. GANs take longer time than VAEs to train.
4. More advanced neural network architectures like Diffusion and Normalizing Flows may even outperform GAN in this task. However, they are computationally expensive and require more hardware resources and time to train.

References

1. https://www.tensorflow.org/tutorials/generative/cvae
2. https://www.kaggle.com/code/bharatadhikari/prototype-gan-restoration-image
3. https://www.geeksforgeeks.org/variational-autoencoders/
4. https://www.geeksforgeeks.org/generative-adversarial-network-gan/